Immunochip Analyses of Epistasis in Rheumatoid Arthritis Confirm Multiple Interactions within MHC and Suggest Novel Non-MHC Epistatic Signals

Wen-Hua Wei, Chia-Yin Loh, Jane Worthington, and Stephen Eyre

ABSTRACT. ***Objective.*** Studying statistical gene-gene interactions (epistasis) has been limited by the difficulties in performance, both statistically and computationally, in large enough sample numbers to gain sufficient power. Three large Immunochip datasets from cohort samples recruited in the United Kingdom, United States, and Sweden with European ancestry were used to examine epistasis in rheumatoid arthritis (RA).

Methods. A full pairwise search was conducted in the UK cohort using a high-throughput tool and the resultant significant epistatic signals were tested for replication in the United States and Swedish cohorts. A forward selection approach was applied to remove redundant signals, while conditioning on the preidentified additive effects.

Results. We detected abundant genome-wide significant ($p < 1.0e-13$) epistatic signals, all within the MHC region. These signals were reduced substantially, but a proportion remained significant ($p < 1.0e-03$) in conditional tests. We identified 11 independent epistatic interactions across the entire MHC, each explaining on average 0.12% of the phenotypic variance, nearly all replicated in both replication cohorts. We also identified non-MHC epistatic interactions between RA susceptible loci *LOC100506023* and *IRF5* with Immunochip-wide significance ($p < 1.1e-08$) and between 2 neighboring single-nucleotide polymorphism near *PTPN22* that were in low linkage disequilibrium with independent interaction ($p < 1.0e-05$). Both non-MHC epistatic interactions were statistically replicated with a similar interaction pattern in the US cohort only.

Conclusion. There are multiple but relatively weak interactions independent of the additive effects in RA and a larger sample number is required to confidently assign additional non-MHC epistasis. (J Rheumatol First Release February 15 2016; doi:10.3899/jrheum.150836)

Key Indexing Terms:
EPISTASIS GENOME-WIDE ASSOCIATION STUDY IMMUNOCHIP
MAJOR HISTOCOMPATIBILITY COMPLEX RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a chronic autoimmune disease affecting about 1% of the population worldwide[1], which if left untreated can lead to progressive synovial joint destruction and subsequent movement disability. Prognosis and treatment at the early stage of the disease are critical to prevent or delay irreparable joint and bone damage[2]. It is therefore important to understand the pathogenesis of RA and the underlying molecular mechanisms to aid development of effective diagnostic tools and treatments.

RA is a complex disease resulting from the interplay of genetic and environmental risk factors. More than 50% of the RA disease risk in twins can be explained by the genetic

From the Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester; National Institute for Health Research (NIHR) Manchester Musculoskeletal Biomedical Research Unit, Central Manchester National Health Service (NHS) Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

Supported by the Higher Education Funding Council for England.

W.H. Wei, PhD, Lecturer in Statistical Genetics, Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester; C.Y. Loh*, MRes, PhD Student, Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester; J. Worthington, PhD, Professor of Chronic Disease Genetics, Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, and NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre; S. Eyre, PhD, Senior Research Fellow on Rheumatological Disorders, Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, and NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre. * These authors contributed equally to this study.*

Address correspondence to Dr. W.H. Wei, Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PT, UK. E-mail: wenhua.wei@manchester.ac.uk

Accepted for publication January 6, 2016.

factors[3]. To date, over 100 RA risk loci have been discovered as a result of genome-wide association studies (GWAS)[4,5]. GWAS primarily consider additive genetic effects and rely on large sample sizes and high-density single-nucleotide polymorphism (SNP) markers to detect loci with moderate or weak effects. Despite the success to date in identifying risk loci for RA, mostly through metaanalysis of multiple cohorts, about 40% of the genetic component of susceptibility remains unaccounted for. One possible explanation for this is the failure to consider nonadditive effects such as dominance[6] and statistical gene-gene interactions (epistasis)[7,8].

So far, analyses of epistasis in human complex traits have been limited by the difficulties, both statistically and computationally, in doing so in large enough sample numbers to gain sufficient power, and as a result only a handful of epistatic interactions have been firmly identified[9]. Such epistatic interactions each carry a moderate effect and thus explain only limited genetic variation of the complex trait studied, but appear essential in modifying the functions of the interacting genes[7,8,9]. In RA, previous studies[10,11] have reported abundant epistatic interactions between SNP mapping to the MHC region without considering the additive effects of the major susceptibility loci within the region. When adjusting for the effects of the well-established HLA-shared epitope (i.e., HLA-DRB1 alleles encoding a 5 amino acid sequence motif)[1,12], the MHC interactions were hugely reduced, but 2 interactions between HLA class II genes remained statistically significant, indicating there exist moderate epistatic interactions independent of the HLA additive effects[13]. Indeed, focusing on only the *HLA-DRB1* locus and using high-resolution imputation in a large combined RA cohort, Lenz, *et al*[14] revealed 7 interactions between *HLA-DRB1* haplotypes that jointly explained an additional 0.5% of the phenotypic variance. Detection of epistasis involving non-MHC loci in RA has been rather challenging[10,13]. An epistatic interaction between *BANK1* and *BLK* in RA was reported from a candidate gene study[15] following the finding of *BANK1-BLK* interaction in systemic lupus erythematosus (SLE) with some protein-protein interaction evidence[16]. In both cases, the reported interaction was just nominally significant (1.0e-02 < p < 5.0e-02), suggesting it might not be detectable at the genome-wide level.

The power of detection of epistasis is a function of the interaction effect size, sample size, and the linkage disequilibrium (LD) between markers and the causal variants at both loci[9]. While interaction effects are probably moderate/weak, increasing the sample size and marker density is the most effective approach to increase power. Besides, examining local interactions between SNP closely located but with low LD (r^2 < 0.2) could increase the power of detection of missing variants and/or functional interactions[17,18]. The use of the Immunochip, a custom array of about 200,000 SNP specifically designed for the fine mapping of a set of immune-mediated disease loci[19], has resulted in the generation of genetic data on much larger numbers of samples for RA and other diseases than was previously achieved through the use of genome-wide arrays and has allowed the refinement of the association signal for multiple autoimmune disease loci[4]. A study of epistasis in primary biliary cirrhosis in a large Immunochip dataset[20] suggests the potential to detect non-MHC interactions.

Here we used 3 large Immunochip datasets of cohort samples recruited in the United Kingdom, United States, and Sweden, to examine epistasis in RA. We performed an exhaustive pairwise search for epistatic signals in the UK cohort and used the other 2 cohorts to test for replication of any significant signals. The biological plausibility of any statistically significant interactions was also investigated.

MATERIALS AND METHODS

Subjects, genotyping, and quality control. We used 3 large cohort samples from the Rheumatoid Arthritis Consortium International for Immunochip that were recruited in the United Kingdom (RACI-UK), United States (RACI-US), and Sweden (RACI-SE, i.e., Swedish Epidemiological Investigation of Rheumatoid Arthritis). The 3 cohorts have been described in detail elsewhere[4,5]. Briefly, all patients with RA fulfilled the 1987 criteria of the American College of Rheumatology[21]; all participants provided written informed consent as approved by a corresponding ethical committee and were genotyped by Immunochip in accordance with Illumina protocols. Only samples with European origin and autosomal SNP were used. SNP genomic locations were in the GRCh38/hg38 version throughout and the MHC was referred to as between 28.7 to 34.0 Mb on chromosome 6.

For each cohort, quality control was conducted using PLINK[22] according to the following criteria: SNP minor allele frequency > 0.01, SNP call rate > 0.95, sample call rate > 0.99, and deviation from Hardy–Weinberg equilibrium p < 1.0e-04. Then the GCTA software[23] was used first to compute the genomic relationship matrix (GRM) and then to remove any related samples through the polygenic prediction procedure using a GRM relatedness threshold of 0.15 recommended for Immunochip[24]. In total, we had 12,266 unrelated samples (3856 cases and 8410 controls, 59.3% women) and 124,453 SNP in the RACI-UK, 4658 samples (2532 cases and 2126 controls, 70.5% women) and 118,678 SNP in the RACI-US, and 4682 samples (2750 cases and 1932 controls, 71.6% women) and 123,315 SNP in the RACI-SE in subsequent analyses.

Statistical analysis. PLINK was used to perform a GWAS in the RACI-UK while fitting sex as covariates in a logistic regression model. The consensus genome-wide significance threshold of 5.0e-08 was applied.

A 2-step approach was used to detect epistasis. First, BiForce[25] was used to perform a full pairwise screening without correction for covariates in the RACI-UK and retained all possible epistatic pairs of SNP with an interaction p value (p_{int}) < 1.0e-05. The retained SNP pairs were adjusted for sex in a full logistic regression model using an R package GenABEL[26], and assessed using the adjusted p_{int} values hereafter. The genotypic model of epistasis was applied in both steps where interaction was tested with 4 degrees of freedom as previously described[27]. A recommended threshold of 1.0e-13[9] was adopted to claim genome-wide significant epistatic SNP pairs. We also adopted an Immunochip-wide 5% significance threshold of 1.1e-08 based on the estimate of 3000 effective number of SNP[24]. In addition, local interactions were extracted if both SNP were located on the same chromosome and within a distance < 1 Mb, with p_{int} < 1.0e-05 and r^2 < 0.2[17].

A forward selection approach was applied iteratively to the MHC genome-wide significant signals to identify first the independent GWAS SNP as the additive background and then independent epistatic SNP pairs conditioning on the additive background. Within each iteration, the most associated SNP or SNP pair (i.e., with the lowest conditional p/p_{int} value) was selected and fixed in the logistic regression model fitting the covariates

as well as any prefixed SNP or SNP pairs. The conditional p or p_{int} values for the remaining SNP or SNP pairs were then calculated. The forward selection process terminated when no more SNP (or SNP pairs) were significant (i.e., conditional $p/p_{int} > 5.0e-02$). The forward selection approach was also taken to identify independent non-MHC SNP pairs when multiple pairs of SNP were likely to be describing the same interaction. Variance explained by the selected epistatic SNP pairs was calculated using a full logistic regression model fitting all the covariates and independent SNP and SNP pairs.

The SNP pairs identified in the RACI-UK were tested for replication in the RACI-US and RACI-SE using the logistic regression model adjusted for sex and considered to be directly replicated if both epistatic SNP had genotype data available and with $p_{int} < 5.0e-02$. Statistically replicated epistatic SNP pairs were further examined for interaction patterns/directions in both cohorts, based on OR and their 95% CI estimated by fitting a logistic regression model that included sex as covariates and each of the 9 possible 2-locus genotypes and using the first homogeneous 2-locus genotype as the reference that by definition had an OR of 1. The OR and CI were plotted using the R package plotrix. The replicated epistatic pairs of SNP were annotated to the closest biologically plausible genes.

RESULTS

The conventional GWAS analysis of the RACI-UK revealed 1746 genome-wide significant SNP ($p < 5.0e-08$), of which the vast majority were within the MHC region. Following the forward selection approach, 20 independent SNP were identified, of which 16 were mapped to the MHC region, 2 to *PTPN22*, 1 to *TNFAIP3*, and the remaining 1 to *ANKRD55* (Supplementary Table 1, available online at jrheum.org).

The full pairwise scan detected 29,921 epistatic pairs of SNP that were genome-wide significant (i.e., $p_{int} < 1.0e-13$), all within the MHC region. Conditioning on the additive background of the 20 independent SNP, 132 of these genome-wide significant SNP pairs had a conditional $p_{int} < 1.0e-03$ (and 2807 additional SNP pairs with a conditional p_{int} between 1.0e-03 and 5.0e-02), indicating that there were indeed epistatic signals within the MHC independent of the additive background. Further forward selection of the 132 SNP pairs resulted in 11 independent epistatic signals representing 3 types of interactions involving either both or 1 or no SNP with genome-wide significant additive effects (Table 1). The 11 independent epistatic pairs of SNP each had a direct replication in both replication cohorts (except for rs805285-rs532098 in the RACI-US), suggesting existence of multiple epistatic interactions across the entire MHC. Conditioning on the additive background that explained 10.8% of the phenotypic variance, the 11 independent pairs of SNP jointly explained an additional 1.7% of the phenotypic variance on the observed scale of which 1.4% by interactions only, suggesting that these interaction effects were not negligible jointly, but weak individually (i.e., on average explained 0.12% of the phenotypic variance).

We found that 50 non-MHC SNP pairs exceeded the Immunochip-wide significance threshold of 1.1e-08 (Supplementary Table 2, available online at jrheum.org), of which only epistatic interactions between rs6425230 (*LOC100506023*) and rs17340646 (near *IRF5*) and between rs17547025 (*LOC101929036*) and rs9806697 (near *SEMA6D*) were directly replicated only in RACI-US and RACI-SE,

Table 1. Independent epistatic SNP pairs within the MHC region adjusted for the preidentified additive background in the RACI-UK and statistical replication in the RACI-US and RACI-SE*.

SNP1	Pos1	Gene1	SNP2	Pos2	Gene2	LD	p_{SNP1_UK}	p_{SNP2_UK}	p_{int_UK}	p_{int-US}	p_{int-SE}	$p_{int-cond}$	$p_{int-final}$
rs805286	31711530	*LY6G6E*	**rs405875**	32247411	Near *NOTCH4*	0.00	2.5e-03	1.5e-36	3.0e-16	2.6e-03	4.2e-05	1.3e-05	1.3e-05
rs154977	32932241	Near *HLA-DMB*	rs1042337	32937203	*HLA-DMB*	0.03	5.2e-22	4.4e-02	1.4e-15	4.4e-05	1.5e-02	1.5e-04	3.9e-04
rs2229094	31572779	*LTA*	**rs2844455**	31901896	*C2*	0.00	1.3e-02	4.4e-21	1.2e-17	7.8e-04	4.8e-04	3.8e-05	1.4e-04
rs3131009	31131055	*PSORS1C1*	**rs9267845**	32225921	Near *NOTCH4*	0.00	5.2e-02	8.1e-08	3.9e-15	5.2e-07	3.2e-04	5.4e-04	1.5e-04
rs9264868	31303824	*HLA-C ~ HLA-B*	rs443198	32222629	*NOTCH4*	0.03	7.4e-01	9.5e-01	1.0e-18	4.4e-02	3.2e-03	1.4e-04	6.1e-04
rs9275653	32718088	Near *HLA-DQA2*	rs2006165	32761010	*HLA-DQB2*	0.11	8.5e-01	6.3e-02	6.2e-29	2.3e-06	1.1e-03	2.2e-04	2.6e-04
rs805285	31711688	Near *LY6G6E*	**rs532098**	32610275	*HLA-DRB1 ~ HLA-DQA1*	0.01	4.9e-02	6.2e-28	5.2e-15	> 0.05	4.2e-02	4.9e-05	4.5e-04
rs9501398	32234820	Near *NOTCH4*	rs3177928	32444658	*HLA-DRA*	0.05	2.3e-01	2.3e-06	4.4e-14	5.5e-07	8.9e-03	5.4e-04	3.3e-03
rs9469220	32690533	Near *HLA-DQB1*	**rs3104407**	32714675	*HLA-DQB1 ~ HLA-DQA2*	0.02	9.3e-03	7.1e-26	5.1e-21	4.4e-21	5.6e-06	7.4e-04	2.6e-03
rs2395471	31272915	Near *HLA-C*	**rs9268658**	32442939	*HLA-DRA*	0.00	2.5e-04	7.8e-39	4.1e-14	8.1e-13	3.5e-04	8.1e-04	5.6e-03
rs5000634	32695787	*HLA-DQB1 ~ HLA-DQA2*	**rs3104407**	32714675	*HLA-DQB1 ~ HLA-DQA2*	0.24	4.4e-51	7.1e-26	5.0e-15	6.6e-04	6.9e-04	7.2e-04	2.9e-02

* The epistatic SNP pairs in the order appeared in the forward selection process. SNP in bold face are genome-wide significant in the GWAS. SNP: single-nucleotide polymorphism; RACI: Rheumatoid Arthritis Consortium International for Immunochip; UK: United Kingdom; US: United States; SE: Sweden; SNP1 (SNP2): the first (second) epistatic SNP; Pos1 (Pos2): the base pair position of SNP1 (SNP2); Gene1 (Gene2): gene mapped by SNP1 (SNP2) – if intergenic annotated as the nearest (with "near") or flanking genes separated by "~"; LD: linkage disequilibrium in r^2; p_{SNP1_UK} (p_{SNP2_UK}): the p values of SNP1 (SNP2) in the logistic regression model fitting genotypes of the 2 SNP and their interactions without adjusting for the MHC additive background; p_{int_UK}: the interaction p values from the pairwise scan of the RACI-UK without adjusting for the additive background; p_{int_US}: the interaction p values from replication tests in the RACI-US without considering the additive background; p_{int_SE}: the interaction p values from replication tests in the RACI-SE without considering the additive background; p_{int_cond}: the interaction p values when conditioning on the additive background in the RACI-UK; p_{int_final}: the interaction p values in the final model fitting the additive background and all the independent pairs in the RACI-UK; > 0.05: interaction p value of the replication test greater than 5.0e-02; GWAS: genome-wide association studies.

Table 2. Non-MHC epistatic pairs of SNP in the RACI-UK and statistical replication in the RACI-US and RACI-SE.

SNP1	Chr1	Pos1	Gene1	SNP2	Chr2	Pos2	Gene2	LD	p_{SNP1_UK}	p_{SNP2_UK}	p_{int_UK}	p_{int-US}	p_{int-SE}	$p_{int-cond}$
rs6425230	1	173418899	*LOC100506023*	rs17340646	7	129082460	Near *IRF5*	0.00	2.8e-01	7.2e-02	4.9e-09	3.6e-02	> 0.05	7.5e-07
rs17547025	12	102932492	*LOC101929036*	rs9806697	15	47778596	Near *SEMA6D*	0.00	5.0e-02	2.1e-01	8.2e-09	> 0.05	2.6e-02	2.8e-08
rs2488457	1	113872746	*PTPN22*	rs17032048	1	113984753	*HIPK1*	0.11	3.0e-11	5.8e-08	6.9e-06	1.4e-05	> 0.05	1.4e-02
rs12711109	4	122091415	Near *KIAA1109*	rs10518402	4	122633635	*IL21-AS1*	0.04	2.4e-02	3.4e-01	2.5e-06	> 0.05	2.8e-02	8.8e-06

SNP: single-nucleotide polymorphism; RACI: Rheumatoid Arthritis Consortium International for Immunochip; UK: United Kingdom; US: United States; SE: Sweden; SNP1 (SNP2): the first (second) epistatic SNP; Chr1 (Chr2): the chromosome of SNP1 (SNP2); Pos1 (Pos2): the base pair position of SNP1 (SNP2); Gene1 (Gene2): gene mapped by SNP1 (SNP2) – if intergenic annotated as the nearest (with "near"); LD: linkage disequilibrium in r^2; p_{SNP1_UK} (p_{SNP2_UK}): the p values of SNP1 (SNP2) in the logistic regression model fitting genotypes of the 2 SNP and their interactions; p_{int_UK}: the interaction p values from the pairwise scan of RACI-UK without adjusting for the additive background; p_{int_US}: the interaction p values from replication tests in the RACI-US without considering the additive background; p_{int_SE}: the interaction p values from replication tests in the RACI-SE without considering the additive background; p_{int_cond}: the interaction p values when conditioning on the additive background in the RACI-UK; > 0.05: interaction p value of the replication test greater than 5.0e-02.

respectively (Table 2). Conditioning on the additive background, the 2 epistatic pairs explained 0.28% and 0.32% of the phenotypic variance in the RACI-UK, respectively, of which the vast majority (i.e., 0.22% and 0.27%, respectively) was contributed by interactions. Nonetheless, only rs6425230 (*LOC10050-6023*)-rs17340646 (near *IRF5*) had a similar interaction pattern in the replication cohort (Figure 1).

We also found 55 non-MHC SNP pairs involved in local interactions (Supplementary Table 3, available online at jrheum.org), of which epistatic interaction between rs2488457 (near *PTPN22*) and rs17032048 (near *HIPK1*) was directly replicated only in the RACI-US, and that between rs12711109 (near *KIAA1109*) and rs10518402 (*IL21-AS1*) was directly replicated only in the RACI-SE (Table 2). Conditioning on the additive background, these 2 epistatic SNP pairs explained 0.11% and 0.23% of phenotypic variance in the RACI-UK, mostly by interactions (i.e., 0.1% and 0.19%), respectively. The rs2488457 (near *PTPN22*)-rs17032048 (near *HIPK1*) pair had a similar interaction pattern in the replication cohort, whereas the rs12711109 (near *KIAA1109*) and rs10518402 (*IL21-AS1*) pair did not share an interaction pattern (Figure 1).

DISCUSSION

Here we present results from a large analysis of epistatic interactions in RA. Compared with previous studies that used a relatively small sample size (e.g., with < 2000 cases and 3000 controls) and a GWAS array[10,11,13], our study had a much increased sample size (i.e., 3856 cases and 8410 controls) and genotype data from the high-density Immunochip, allowing the identification of many more genome-wide significant interactions within the MHC region. Conditioning on the additive background, we confirmed that most of these interaction signals mirrored the strong HLA additive effects, but in addition there were 11 independent interactions each with a moderate/weak effect that were statistically replicated in 2 independent cohorts. These results indicate that, in addition to the recently confirmed epistasis within the *HLA-DRB1* locus[14], the entire MHC region contains an increased level of complexity, with additional associations to disease risk, over and above the multiple independent signals previously reported. Nevertheless, caution is recommended in interpreting the 11 independent epistatic pairs given that the resolution provided by Immunochip SNP is insufficient to identify the MHC diversity and the additive background was derived imperfectly. Further work is needed to finely map the MHC epistatic signals based on high-resolution imputation and large samples as demonstrated elsewhere[14].

The large Immunochip cohort RACI-UK also enabled us to detect non-MHC epistatic interactions (Table 2). Based on the Immunochip-wide significance threshold, we identified the rs6425230 (*LOC100506023*)-rs17340646 (near *IRF5*) epistatic pair that was directly replicated in the RACI-US (Table 2) and with a similar interaction pattern (Figure 1A). Both *LOC100506023* and *IRF5* are known RA-susceptible loci identified from a GWAS meta-analysis[5]. Interestingly, *IRF5* was previously found to be an interacting locus with moderate interactions in autoimmune diseases SLE[28,29,30] and primary biliary cirrhosis[20]. From the local interactions in the RACI-UK (Supplementary Table 3, available online at jrheum.org), we identified the rs2488457 (near *PTPN22*)-rs17032048 (near *HIPK1*) epistatic pair that was also directly replicated in the RACI-US with a similar interaction pattern (Table 2 and Figure 1C). We showed that this local interaction was greatly reduced but remained nominally significant when conditioning on the predefined additive background including the well-known *PTPN22* variant marked by rs2476601 (Supplementary Table 1, available online at jrheum.org). While the biological function of the epistatic pair is unclear, the result at least suggested that there could be independent epistasis involved in regulating *PTPN22*, although the epistasis was relatively weak and not directly replicated in the second replication cohort RACI-SE.

Nonetheless, we reported only SNP interactions that were directly replicated (i.e., both SNP are genotyped and are interacting) with a similar interaction pattern in an independent cohort. As a result of this stringent criteria, we could have

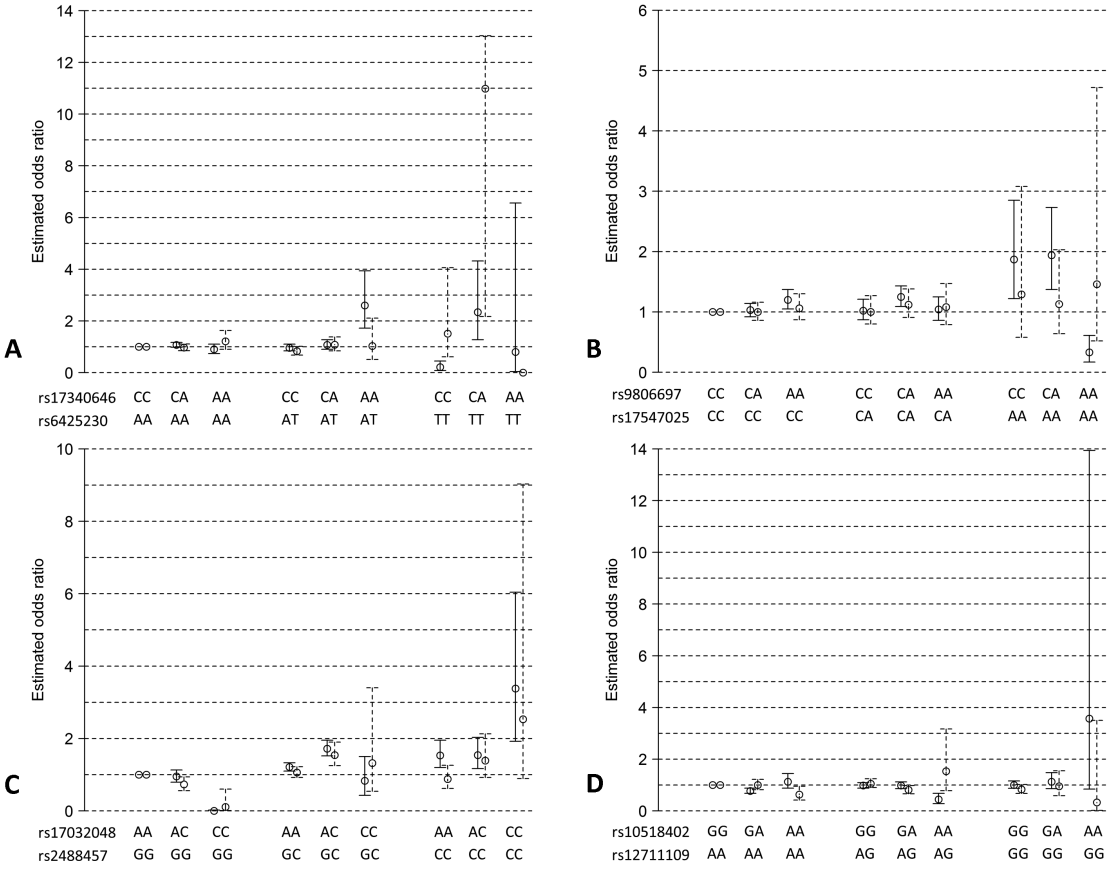

Figure 1. Statistical interaction patterns of 4 directly replicated epistatic pairs of SNP based on or estimated for each of the 9 joint SNP genotypes in the discovery (solid line) and replication (dash line) cohorts. A. rs6425230 (*LOC100506023*)-rs17340646 (near *IRF5*) replicated in the RACI-US. B. rs17547025 (*LOC101929036*)-rs9806697 (near *SEMA6D*) replicated in the RACI-SE. C. rs2488457 (near *PTPN22*)-rs17032048 (near *HIPK1*) replicated in the RACI-US. D. rs12711109 (near *KIAA1109*)-rs10518402 (*IL21-AS1*) replicated in the RACI-SE. Bars represent 95% CI calculated from the standard errors of the estimates of the genotype variables in the logistic regression model used to calculate genotypic OR. SNP: single-nucleotide polymorphism; RACI: Rheumatoid Arthritis Consortium International for Immunochip; US: United States; SE: Sweden.

failed to detect additional non-MHC interactions that are not directly but indirectly replicated through proxies of the epistatic SNP. Besides, the sample size in either the RACI-US or RACI-SE was about 50% smaller than that in the discovery cohort and thus could have limited replication results. Further, special cohort structures, e.g., matched cases and controls used only in the RACI-SE[4], could influence calculation OR and subsequently interaction patterns (Figure 1).

We also tested statistical interactions between the genome-wide significant GWAS SNP in the RACI-UK, but found no significant epistasis beyond the MHC region, which is in agreement with the results by Liu, *et al*[13] that used an allelic model instead (i.e., concerning only SNP additive effects and additive-additive interaction). However, the epistatic interaction between *HLA-DQA2* and *HLA-DQB2* reported by Liu, *et al*[13] was only marginally significant (i.e., p_{int} = 1.9e-02 for rs9275390-rs10807113 in the genotypic model) when adjusting for the additive background in the RACI-UK and thus was not among the 11 independent MHC epistatic signals (Table 1). This discrepancy highlights the limitation of the allelic model in studying epistasis and the importance of applying proper adjustment of preidentified additive effects. In addition, we found that the reported epistatic interactions between *ERAP1* and *HLA-C* in psoriasis[8] and between *ERAP1* and *HLA-B*51* in Behçet disease[31] did not reach the nominal significance level in the RACI-UK (the *ERAP1* and *HLA-B27* interaction in ankylosing spondylitis[7] was not tested because of unavailable genotype data of the epistatic SNP rs4349859), suggesting RA may have different regulatory mechanisms.

Our results indicate that epistasis may exist in regulating complex diseases such as RA, but is unlikely to explain much genetic variance (i.e., with small effect sizes) and hence requires a large sample size to detect conclusively. These results provide fresh evidence to support the theoretical prediction of low amounts of epistatic variance in complex traits in outbred populations[32] that could still be important to understand the underlying regulatory mechanisms. However,

our study does not provide a genome-wide view of epistasis in RA because of the use of Immunochip, and the results may be slightly conservative because of the stringent replication criteria and the inclusion of all patients with RA, irrespective of disease heterogeneity[1].

Taking advantage of increased samples and high density SNP offered by Immunochip, we confirmed the existence of epistatic interactions with the MHC region and suggested novel non-MHC interactions with moderate interaction effects but plausible biological interests. Further work is required to validate the identified epistatic interactions functionally.

ACKNOWLEDGMENT

We thank the Wellcome Trust Sanger Institute Genotyping Facility for genotyping of the UK Rheumatoid Arthritis Genetics samples funded by the Arthritis Research UK grant reference number 17552 and the Manchester Biomedical Research Centre. Paul Gilbert prepared the UK samples. We acknowledge the use of DNA from the UK Blood Services collection of Common Controls (UKBS-CC collection), funded by Wellcome Trust grant 076113/C/04/Z and by the US National Institutes of Health (NIH) research program grant to the National Health Service Blood and Transplant (RP-PG-0310-1002). We acknowledge the use of DNA from the British 1958 Birth Cohort collection, funded by the UK Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. The North American Rheumatoid Arthritis Consortium and analysis of other US patient and control collections at the Feinstein Institute were supported by the US NIH grants RO1-AR-4-4422, NO1-AR-2-2263; NO1-AR1-2256, RO1 AI068759, and RC2AR059092-01, with additional support from the Eileen Ludwig Greenland Center for Rheumatoid Arthritis and the family of Robert S. Boas.

ONLINE SUPPLEMENT

Supplementary data for this article are available online at jrheum.org.

REFERENCES

1. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol 2013;9:141-53.
2. Quinn MA, Cox S. The evidence for early intervention. Rheum Dis Clin North Am 2005;31:575-89.
3. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43:30-7.
4. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al; Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; Wellcome Trust Case Control Consortium. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44:1336-40.
5. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al; RACI consortium; GARNET consortium Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376-81.
6. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al; International Inflammatory Bowel Disease Genetics Consortium; Australia and New Zealand IBDGC; Belgium IBD Genetics Consortium; Italian Group for IBD Genetic Consortium; NIDDK Inflammatory Bowel Disease Genetics Consortium; United Kingdom IBDGC; Wellcome Trust Case Control Consortium; Quebec IBD Genetics Consortium. High-density mapping of the MHC identifies a shared function for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet 2015;47:172-9.
7. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al ; Spondyloarthritis Research Consortium of Canada (SPARCC); Australo-Anglo-American Spondyloarthritis Consortium (TASC); Wellcome Trust Case Control Consortium 2 (WTCCC2). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet 2011;43:761-7.
8. Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2, Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet 2010;42:985-90.
9. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet 2014;15:722-33.
10. Lippert C, Listgarten J, Davidson RI, Baxter S, Poon H, Kadie CM, et al. An exhaustive epistatic SNP association analysis on expanded Wellcome Trust Data. Sci Rep 2013;3:1099.
11. Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NL, et al. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. Am J Hum Genet 2010;87:325-40.
12. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2012;44:291-6.
13. Liu C, Ackerman HH, Carulli JP. A genome-wide screen of gene-gene interactions for rheumatoid arthritis susceptibility. Human Genet 2011;129:473-85.
14. Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, Eyre S, et al. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nat Genet 2015;47:1085-90.
15. Génin E, Coustet B, Allanore Y, Ito I, Teruel M, Constantin A, et al. Epistatic interaction between BANK1 and BLK in rheumatoid arthritis: results from a large trans-ethnic meta-analysis. PLoS One 2013;8:e61044.
16. Castillejo-López C, Delgado-Vega AM, Wojcik J, Kozyrev SV, Thavathiru E, Wu YY, et al. Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. Ann Rheum Dis 2012;71:136-42.
17. Wei W, Gyenesei A, Semple CA, Haley CS. Properties of local interactions and their potential value in complementing genome-wide association studies. PLoS One 2013;8:e71203.
18. Wei WH, Guo Y, Kindt AS, Merriman TR, Semple CA, Wang K, et al. Abundant local interactions in the 4p16.1 region suggest functional mechanisms underlying SLC2A9 associations with human serum uric acid. Hum Mol Genet 2014;23:5061-8.
19. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther 2011;13:101.
20. Juran BD, Hirschfield GM, Invernizzi P, Atkinson EJ, Li Y, Xie G, et al; Italian PBC Genetics Study Group Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. Hum Mol Genet 2012;21:5209-21.
21. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559-75.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011; 88:76-82.

24. Chen GB, Lee SH, Brion MJ, Montgomery GW, Wray NR, Radford-Smith GL, et al; International IBD Genetics Consortium. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Hum Mol Genet 2014;23:4710-20.
25. Gyenesei A, Moody J, Laiho A, Semple CA, Haley CS, Wei WH. BiForce Toolbox: powerful high-throughput computational analysis of gene-gene interactions in genome-wide association studies. Nucleic Acids Res 2012;40:W628-32.
26. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23:1294-6.
27. Gyenesei A, Moody J, Semple CA, Haley CS, Wei WH. High-throughput analysis of epistasis in genome-wide association studies with BiForce. Bioinformatics 2012;28:1957-64.
28. Dang J, Shan S, Li J, Zhao H, Xin Q, Liu Y, et al. Gene-gene interactions of IRF5, STAT4, IKZF1 and ETS1 in systemic lupus erythematosus. Tissue Antigens 2014;83:401-8.
29. Hellquist A, Jarvinen TM, Koskenmies S, Zucchelli M, Orsmark-Pietras C, Berglind L, et al. Evidence for genetic association and interaction between the TYK2 and IRF5 genes in systemic lupus erythematosus. J Rheumatol 2009;36:1631-8.
30. Tang L, Wan P, Wang Y, Pan J, Wang Y, Chen B. Genetic association and interaction between the IRF5 and TYK2 genes and systemic lupus erythematosus in the Han Chinese population. Inflamm Res 2015;64:817-24.
31. Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, et al. Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet 2013;45:202-7.
32. Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. Genetics 2014;198:355-67.